CLAIMS

What is claimed is:

1. A machine-implemented method comprising:

extracting portions from time-domain speech segments;

creating feature vectors that represent the portions in a vector space, the feature vectors incorporating phase information of the portions; and

determining a distance between the feature vectors in the vector space.

2. The machine-implemented method of claim 1, wherein creating feature vectors comprises:

constructing a matrix W from the portions; and

decomposing the matrix W.

3. The machine-implemented method of claim 2, wherein decomposing the matrix W comprises extracting global boundary-centric features from the portions.

4. The machine-implemented method of claim 2, wherein the speech segments each include a segment boundary within a phoneme.

5. The machine-implemented method of claim 4, wherein the speech segments each include at least one diphone.

6. The machine-implemented method of claim 5, wherein the portions include at least one pitch period.

7. The machine-implemented method of claim 6, wherein decomposing the matrix W comprises performing a pitch synchronous singular value analysis on the pitch periods of the time-domain segments.

8. The machine-implemented method of claim 6, wherein the matrix W is a $2KM \times N$ matrix represented by

$$W = U \Sigma V^T$$

where K is the number of pitch periods near the segment boundary extracted from each segment, N is the maximum number of samples among the pitch periods, M is the number of segments in the voice table having a segment boundary within the phoneme, U is the $2KM \times R$ left singular matrix with row vectors u_i ($1 \leq i \leq 2KM$), Σ is the $R \times R$ diagonal matrix of singular values $s_1 \geq s_2 \geq \ldots \geq s_R > 0$, V is the $N \times R$ right singular matrix with row vectors v_j ($1 \leq j \leq N$), $R \ll 2KM$, and ^{T} denotes matrix transposition, wherein decomposing the matrix W comprises performing a singular value decomposition of W.

9. The machine-implemented method of claim 8, wherein the pitch periods are zero padded to N samples.

10. The machine-implemented method of claim 9, wherein a feature vector $\bar{u}_i$ is calculated as

$$\bar{u}_i = u_i \Sigma$$

where u_i is a row vector associated with a pitch period i, and Σ is the singular diagonal matrix.

11. The machine-implemented method of claim 10, wherein the distance between two feature vectors is determined by a metric comprising the cosine of the angle between the two feature vectors.

12. The machine-implemented method of claim 11, wherein the metric comprises a closeness measure, C, between two feature vectors, $\bar{u}_k$ and $\bar{u}_l$, wherein C is calculated as

$$C(\bar{u}_k, \bar{u}_l) = \cos(u_k\Sigma, u_l\Sigma) = \frac{u_k \Sigma^2 u_l^T}{\| u_k \Sigma \| \; \| u_l \Sigma \|}$$

for any $1 \leq k, l \leq 2KM$.

13. The machine-implemented method of claim 12, wherein a difference $d(S_1,S_2)$ between two segments in the voice table, S_1 and S_2, is calculated as

$$d(S_1,S_2) = d_0(p_1, q_1) = 1 - C(\bar{u}_{p1}, \bar{u}_{q1})$$

where d_0 is the distance between pitch periods p_1 and q_1, p_1 is the last pitch period of S_1, q_1 is the first pitch period of S_2, $\bar{u}_{p1}$ is a feature vector associated with pitch period p_1, and $\bar{u}_{q1}$ is a feature vector associated with pitch period q_1.

14. The machine-implemented method of claim 13, wherein the calculation for the difference between two segments in the voice table, S_1 and S_2, is expanded to include a plurality of pitch periods from each segment.

15. The machine-implemented method of claim 13, wherein the difference between two segments in the voice table, S_1 and S_2, is associated with a discontinuity between S_1 and S_2.

16. The machine-implemented method of claim 12, wherein a difference $d(S_1,S_2)$ between two segments in the voice table, S_1 and S_2, is calculated as

$$d(S_1,S_2) = \left| \frac{d_0(p_1, q_1) - d_0(p_1, \overline{p}_1) + d_0(q_1, \overline{q}_1)}{2} \right| = \left| \frac{C(\overline{u}_{p1}, \overline{u}_{\overline{p}1}) + C(\overline{u}_{q1}, \overline{u}_{\overline{q}1}) - C(\overline{u}_{p1}, \overline{u}_{q1})}{2} \right|$$

where d_0 is the distance between pitch periods, p_1 is the last pitch period of S_1, $\overline{p}_1$ is the first pitch period of a segment contiguous to S_1, q_1 is the first pitch period of S_2, $\overline{q}_1$ is the last pitch period of a segment contiguous to S_2, $\overline{u}_{p1}$ is a feature vector associated with pitch period p_1, $\overline{u}_{q1}$ is a feature vector associated with pitch period q_1, $\overline{u}_{\overline{p}1}$ is a feature vector associated with pitch period $\overline{p}_1$, and $\overline{u}_{\overline{q}1}$ is a feature vector associated with pitch period $\overline{q}_1$.

17. The machine-implemented method of claim 2, further comprising associating the distance between the feature vectors with speech segments in the voice table.

18. The machine-implemented method of claim 17, further comprising:

selecting speech segments from the voice table based on the distance between the feature vectors.

19. The machine-implemented method of claim 5, wherein the portions include centered pitch periods.

20. The machine-implemented method of claim 19, wherein the matrix W is a $(2(K\text{-}1)+1)M \times N$ matrix represented by

$$W = U \Sigma V^T$$

where K-1 is the number of centered pitch periods near the segment boundary extracted from each segment, N is the maximum number of samples among the centered pitch periods, M is the number of segments in the voice table having a segment boundary within the phoneme, U is the $(2(K\text{-}1)+1)M \times R$ left singular matrix with row vectors u_i $(1 \leq i \leq (2(K\text{-}1)+1)M)$, Σ is the $R \times R$ diagonal matrix of singular values $s_1 \geq s_2 \geq \ldots \geq s_R > 0$, V is the $N \times R$ right singular matrix with row vectors v_j $(1 \leq j \leq N)$, $R \ll (2(K\text{-}1)+1)M)$, and ^{T} denotes matrix transposition, wherein decomposing the matrix W comprises performing a singular value decomposition of W.

21. The machine-implemented method of claim 20, wherein the centered pitch periods are symmetrically zero padded to N samples.

22. The machine-implemented method of claim 21, wherein a feature vector $\bar{u}_i$ is calculated as

$$\bar{u}_i = u_i \Sigma$$

where u_i is a row vector associated with a centered pitch period i, and Σ is the singular diagonal matrix.

23. The machine-implemented method of claim 22, wherein the distance between two feature vectors is determined by a metric comprising a closeness measure, C, between two feature vectors, $\bar{u}_k$ and $\bar{u}_l$, wherein C is calculated as

$$C(\bar{u}_k, \bar{u}_l) = \cos(u_k\Sigma, u_l\Sigma) = \frac{u_k \Sigma^2 u_l^T}{\| u_k \Sigma \| \; \| u_l \Sigma \|}$$

for any $1 \leq k, l \leq (2(K\text{-}1)+1)M$.

24. The machine-implemented method of claim 23, wherein a difference $d(S_1,S_2)$ between two segments in the voice table, S_1 and S_2, is calculated as

$$d(S_1,S_2) = C(u_{\pi_{-1}}, u_{\delta_0}) + C(u_{\delta_0}, u_{\sigma_1}) - C(u_{\pi_{-1}}, u_{\pi_0}) - C(u_{\sigma_0}, u_{\sigma_1})$$

where $u_{\pi_{-1}}$ is a feature vector associated with a centered pitch period π_{-1}, u_{δ_0} is a feature vector associated with a centered pitch period δ_0, u_{σ_1} is a feature vector associated with a centered pitch period σ_1, u_{π_0} is a feature vector associated with a centered pitch period π_0, and u_{σ_0} is a feature vector associated with a centered pitch period σ_0.

25. A machine-readable medium having instructions to cause a machine to perform a machine-implemented method comprising:

extracting portions from time-domain speech segments;

creating feature vectors that represent the portions in a vector space, the feature vectors incorporating phase information of the portions; and

determining a distance between the feature vectors in the vector space.

26. The machine-readable medium of claim 25, wherein creating feature vectors comprises:

constructing a matrix W from the portions; and

decomposing the matrix W.

27. The machine-readable medium of claim 26, wherein decomposing the matrix W comprises extracting global boundary-centric features from the portions.

28. The machine-readable medium of claim 26, wherein the speech segments each include a segment boundary within a phoneme.

29. The machine-readable medium of claim 28, wherein the speech segments each include at least one diphone.

30. The machine-readable medium of claim 29, wherein the portions include at least one pitch period.

31. The machine-readable medium of claim 30, wherein decomposing the matrix W comprises performing a pitch synchronous singular value analysis on the pitch periods of the time-domain segments.

32. The machine-readable medium of claim 30, wherein the matrix W is a $2KM \times N$ matrix represented by

$$W = U \Sigma V^T$$

where K is the number of pitch periods near the segment boundary extracted from each segment, N is the maximum number of samples among the pitch periods, M is the number of segments in the voice table having a segment boundary within the phoneme, U is the $2KM \times R$ left singular matrix with row vectors u_i ($1 \leq i \leq 2KM$), Σ is the $R \times R$ diagonal matrix of singular values $s_1 \geq s_2 \geq \ldots \geq s_R > 0$, V is the $N \times R$ right singular matrix with row vectors v_j ($1 \leq j \leq N$), $R << 2KM$, and ^{T} denotes matrix transposition, wherein decomposing the matrix W comprises performing a singular value decomposition of W.

33. The machine-readable medium of claim 32, wherein the pitch periods are zero padded to N samples.

34. The machine-readable medium of claim 33, wherein a feature vector $\bar{u}_i$ is calculated as

$$\bar{u}_i = u_i \Sigma$$

where u_i is a row vector associated with a pitch period i, and Σ is the singular diagonal matrix.

35. The machine-readable medium of claim 34, wherein the distance between two feature vectors is determined by a metric comprising the cosine of the angle between the two feature vectors.

36. The machine-readable medium of claim 35, wherein the metric comprises a closeness measure, C, between two feature vectors, $\bar{u}_k$ and $\bar{u}_l$, wherein C is calculated as

$$C(\bar{u}_k, \bar{u}_l) = \cos(u_k\Sigma, u_l\Sigma) = \frac{u_k \Sigma^2 u_l^T}{\| u_k \Sigma \| \, \| u_l \Sigma \|}$$

for any $1 \leq k, l \leq 2KM$.

37. The machine-readable medium of claim 36, wherein a difference $d(S_1,S_2)$ between two segments in the voice table, S_1 and S_2, is calculated as

$$d(S_1,S_2) = d_0(p_1, q_1) = 1 - C(\bar{u}_{p1}, \bar{u}_{q1})$$

where d_0 is the distance between pitch periods p_1 and q_1, p_1 is the last pitch period of S_1, q_1 is the first pitch period of S_2, $\bar{u}_{p1}$ is a feature vector associated with pitch period p_1, and $\bar{u}_{q1}$ is a feature vector associated with pitch period q_1.

38. The machine-readable medium of claim 37, wherein the calculation for the difference between two segments in the voice table, S_1 and S_2, is expanded to include a plurality of pitch periods from each segment.

39. The machine-readable medium of claim 37, wherein the difference between two segments in the voice table, S_1 and S_2, is associated with a discontinuity between S_1 and S_2.

40. The machine-readable medium of claim 36, wherein a difference $d(S_1,S_2)$ between two segments in the voice table, S_1 and S_2, is calculated as

$$d(S_1,S_2) = \left| \frac{d_0(p_1, q_1) - d_0(p_1, \overline{p}_1) + d_0(q_1, \overline{q}_1)}{2} \right| = \left| \frac{C(\overline{u}_{p1}, \overline{u}_{\overline{p}1}) + C(\overline{u}_{q1}, \overline{u}_{\overline{q}1}) - C(\overline{u}_{p1}, \overline{u}_{q1})}{2} \right|$$

where d_0 is the distance between pitch periods, p_1 is the last pitch period of S_1, $\overline{p}_1$ is the first pitch period of a segment contiguous to S_1, q_1 is the first pitch period of S_2, $\overline{q}_1$ is the last pitch period of a segment contiguous to S_2, $\overline{u}_{p1}$ is a feature vector associated with pitch period p_1, $\overline{u}_{q1}$ is a feature vector associated with pitch period q_1, $\overline{u}_{\overline{p}1}$ is a feature vector associated with pitch period $\overline{p}_1$, and $\overline{u}_{\overline{q}1}$ is a feature vector associated with pitch period $\overline{q}_1$.

41. The machine-readable medium of claim 26, wherein the method further comprises associating the distance between the feature vectors with speech segments in the voice table.

42. The machine-readable medium of claim 41, wherein the method further comprises:

selecting speech segments from the voice table based on the distance between the feature vectors.

43. The machine-readable medium of claim 29, wherein the portions include centered pitch periods.

44. The machine-readable medium of claim 43, wherein the matrix W is a $(2(K-1)+1)M \times N$ matrix represented by

$$W = U \Sigma V^T$$

where K-1 is the number of centered pitch periods near the segment boundary extracted from each segment, N is the maximum number of samples among the centered pitch periods, M is the number of segments in the voice table having a segment boundary within the phoneme, U is the $(2(K-1)+1)M \times R$ left singular matrix with row vectors u_i $(1 \le i \le (2(K-1)+1)M)$, Σ is the $R \times R$ diagonal matrix of singular values $s_1 \ge s_2 \ge \ldots \ge s_R > 0$, V is the $N \times R$ right singular matrix with row vectors v_j $(1 \le j \le N)$, $R \ll (2(K-1)+1)M)$, and ^{T} denotes matrix transposition, wherein decomposing the matrix W comprises performing a singular value decomposition of W.

45. The machine-readable medium of claim 44, wherein the centered pitch periods are symmetrically zero padded to N samples.

46. The machine-readable medium of claim 45, wherein a feature vector $\bar{u}_i$ is calculated as

$$\bar{u}_i = u_i \Sigma$$

where u_i is a row vector associated with a centered pitch period i, and Σ is the singular diagonal matrix.

47. The machine-readable medium of claim 46, wherein the distance between two feature vectors is determined by a metric comprising a closeness measure, C, between two feature vectors, $\bar{u}_k$ and $\bar{u}_l$, wherein C is calculated as

$$C(\bar{u}_k, \bar{u}_l) = \cos(u_k\Sigma, u_l\Sigma) = \frac{u_k \Sigma^2 u_l^T}{\| u_k \Sigma \| \; \| u_l \Sigma \|}$$

for any $1 \leq k, l \leq (2(K\text{-}1)+1)M$.

48. The machine-readable medium of claim 47, wherein a difference $d(S_1,S_2)$ between two segments in the voice table, S_1 and S_2, is calculated as

$$d(S_1,S_2) = C(u_{\pi_{-1}}, u_{\delta_0}) + C(u_{\delta_0}, u_{\sigma_1}) - C(u_{\pi_{-1}}, u_{\pi_0}) - C(u_{\sigma_0}, u_{\sigma_1})$$

where $u_{\pi_{-1}}$ is a feature vector associated with a centered pitch period π_{-1}, u_{δ_0} is a feature vector associated with a centered pitch period δ_0, u_{σ_1} is a feature vector associated with a centered pitch period σ_1, u_{π_0} is a feature vector associated with a centered pitch period π_0, and u_{σ_0} is a feature vector associated with a centered pitch period σ_0.

49. An apparatus comprising:

means for extracting portions from time-domain speech segments;

means for creating feature vectors that represent the portions in a vector space, the feature vectors incorporating phase information of the portions; and

means for determining a distance between the feature vectors in the vector space.

50. The apparatus of claim 49, wherein creating feature vectors comprises:

means for constructing a matrix W from the portions; and

means for decomposing the matrix W.

51. The apparatus of claim 50, wherein the means for decomposing the matrix W comprises means for extracting global boundary-centric features from the portions.

52. The apparatus of claim 50, wherein the speech segments each include a segment boundary within a phoneme.

53. The apparatus of claim 52, wherein the speech segments each include at least one diphone.

54. The apparatus of claim 53, wherein the portions include at least one pitch period.

55. The apparatus of claim 54, wherein the means for decomposing the matrix W comprises means for performing a pitch synchronous singular value analysis on the pitch periods of the time-domain segments.

56. The apparatus of claim 54, wherein the matrix W is a $2KM \times N$ matrix represented by

$$W = U \Sigma V^T$$

where K is the number of pitch periods near the segment boundary extracted from each segment, N is the maximum number of samples among the pitch periods, M is the number of segments in the voice table having a segment boundary within the phoneme, U is the $2KM \times R$ left singular matrix with row vectors u_i ($1 \leq i \leq 2KM$), Σ is the $R \times R$ diagonal matrix of singular values $s_1 \geq s_2 \geq \ldots \geq s_R > 0$, V is the $N \times R$ right singular matrix with row vectors v_j ($1 \leq j \leq N$), $R << 2KM$, and ^{T} denotes matrix transposition, wherein decomposing the matrix W comprises performing a singular value decomposition of W.

57. The apparatus of claim 56, wherein the pitch periods are zero padded to N samples.

58. The apparatus of claim 57, wherein a feature vector $\bar{u}_i$ is calculated as

$$\bar{u}_i = u_i \Sigma$$

where u_i is a row vector associated with a pitch period i, and Σ is the singular diagonal matrix.

59. The apparatus of claim 58, wherein the distance between two feature vectors is determined by a metric comprising the cosine of the angle between the two feature vectors.

60. The apparatus of claim 59, wherein the metric comprises a closeness measure, C, between two feature vectors, $\bar{u}_k$ and $\bar{u}_l$, wherein C is calculated as

$$C(\bar{u}_k, \bar{u}_l) = \cos(u_k\Sigma, u_l\Sigma) = \frac{u_k \Sigma^2 u_l{}^T}{\| u_k \Sigma \| \ \| u_l \Sigma \|}$$

for any $1 \leq k, l \leq 2KM$.

61. The apparatus of claim 60, wherein a difference $d(S_1,S_2)$ between two segments in the voice table, S_1 and S_2, is calculated as

$$d(S_1,S_2) = d_0(p_1, q_1) = 1 - C(\bar{u}_{p1}, \bar{u}_{q1})$$

where d_0 is the distance between pitch periods p_1 and q_1, p_1 is the last pitch period of S_1, q_1 is the first pitch period of S_2, $\bar{u}_{p1}$ is a feature vector associated with pitch period p_1, and $\bar{u}_{q1}$ is a feature vector associated with pitch period q_1.

62. The apparatus of claim 61, wherein the calculation for the difference between two segments in the voice table, S_1 and S_2, is expanded to include a plurality of pitch periods from each segment.

63. The apparatus of claim 61, wherein the difference between two segments in the voice table, S_1 and S_2, is associated with a discontinuity between S_1 and S_2.

64. The apparatus of claim 60, wherein a difference $d(S_1,S_2)$ between two segments in the voice table, S_1 and S_2, is calculated as

$$d(S_1,S_2) = \left| \frac{d_0(p_1, q_1) - d_0(p_1, \bar{p}_1) + d_0(q_1, \bar{q}_1)}{2} \right| = \left| \frac{C(\bar{u}_{p1}, \bar{u}_{\bar{p}1}) + C(\bar{u}_{q1}, \bar{u}_{\bar{q}1}) - C(\bar{u}_{p1}, \bar{u}_{q1})}{2} \right|$$

where d_0 is the distance between pitch periods, p_1 is the last pitch period of S_1, $\bar{p}_1$ is the first pitch period of a segment contiguous to S_1, q_1 is the first pitch period of S_2, $\bar{q}_1$ is the last pitch period of a segment contiguous to S_2, $\bar{u}_{p1}$ is a feature vector associated with pitch period p_1, $\bar{u}_{q1}$ is a feature vector associated with pitch period q_1, $\bar{u}_{\bar{p}1}$ is a feature vector associated with pitch period $\bar{p}_1$, and $\bar{u}_{\bar{q}1}$ is a feature vector associated with pitch period $\bar{q}_1$.

65. The apparatus of claim 50, further comprising means for associating the distance between the feature vectors with speech segments in the voice table.

66. The apparatus of claim 65, further comprising:

means for selecting speech segments from the voice table based on the distance between the feature vectors.

67. The apparatus of claim 53, wherein the portions include centered pitch periods.

68. The apparatus of claim 67, wherein the matrix W is a $(2(K\text{-}1)+1)M \times N$ matrix represented by

$$W = U \Sigma V^T$$

where K-1 is the number of centered pitch periods near the segment boundary extracted from each segment, N is the maximum number of samples among the centered pitch periods, M is the number of segments in the voice table having a segment boundary within the phoneme, U is the $(2(K\text{-}1)+1)M \times R$ left singular matrix with row vectors u_i $(1 \leq i \leq (2(K\text{-}1)+1)M)$, Σ is the $R \times R$ diagonal matrix of singular values $s_1 \geq s_2 \geq \ldots \geq s_R > 0$, V is the $N \times R$ right singular matrix with row vectors v_j $(1 \leq j \leq N)$, $R \ll (2(K\text{-}1)+1)M)$, and ^{T} denotes matrix transposition, wherein decomposing the matrix W comprises performing a singular value decomposition of W.

69. The apparatus of claim 68, wherein the centered pitch periods are symmetrically zero padded to N samples.

70. The apparatus of claim 69, wherein a feature vector $\bar{u}_i$ is calculated as

$$\bar{u}_i = u_i \Sigma$$

where u_i is a row vector associated with a centered pitch period i, and Σ is the singular diagonal matrix.

71. The apparatus of claim 70, wherein the distance between two feature vectors is determined by a metric comprising a closeness measure, C, between two feature vectors, $\bar{u}_k$ and $\bar{u}_l$, wherein C is calculated as

$$C(\bar{u}_k, \bar{u}_l) = \cos(u_k\Sigma, u_l\Sigma) = \frac{u_k \Sigma^2 u_l^T}{\| u_k \Sigma \| \; \| u_l \Sigma \|}$$

for any $1 \leq k, l \leq (2(K\text{-}1)+1)M$.

72. The apparatus of claim 71, wherein a difference $d(S_1,S_2)$ between two segments in the voice table, S_1 and S_2, is calculated as

$$d(S_1,S_2) = C(u\pi_{-1}, u\delta_0) + C(u\delta_0, u\sigma_1) - C(u\pi_{-1}, u\pi_0) - C(u\sigma_0, u\sigma_1)$$

where $u\pi_{-1}$ is a feature vector associated with a centered pitch period π_{-1}, $u\delta_0$ is a feature vector associated with a centered pitch period δ_0, $u\sigma_1$ is a feature vector associated with a centered pitch period σ_1, $u\pi_0$ is a feature vector associated with a centered pitch period π_0, and $u\sigma_0$ is a feature vector associated with a centered pitch period σ_0.

73. A system comprising:

a processing unit coupled to a memory through a bus; and

a process executed from the memory by the processing unit to cause the processing unit to extract portions from time-domain speech segments, create feature vectors that represent the portions in a vector space, the feature vectors incorporating phase information of the portions, and determine a distance between the feature vectors in the vector space.

74. The system of claim 73, wherein the process further causes the processing unit, when creating feature vectors, to construct a matrix *W* from the portions, and decompose the matrix *W*.

75. The system of claim 74, wherein the process further causes the processing unit, when decomposing the matrix *W*, to extract global boundary-centric features from the portions.

76. The system of claim 74, wherein the speech segments each include a segment boundary within a phoneme.

77. The system of claim 76, wherein the speech segments each include at least one diphone.

78. The system of claim 77, wherein the portions include at least one pitch period.

79. The system of claim 78, wherein the process further causes the processing unit, when decomposing the matrix *W*, to perform a pitch synchronous singular value analysis on the pitch periods of the time-domain segments.

80. The system of claim 78, wherein the matrix W is a $2KM \times N$ matrix represented by

$$W = U \Sigma V^T$$

where K is the number of pitch periods near the segment boundary extracted from each segment, N is the maximum number of samples among the pitch periods, M is the number of segments in the voice table having a segment boundary within the phoneme, U is the $2KM \times R$ left singular matrix with row vectors u_i ($1 \leq i \leq 2KM$), Σ is the $R \times R$ diagonal matrix of singular values $s_1 \geq s_2 \geq \ldots \geq s_R > 0$, V is the $N \times R$ right singular matrix with row vectors v_j ($1 \leq j \leq N$), $R << 2KM$, and ^{T} denotes matrix transposition, wherein decomposing the matrix W comprises performing a singular value decomposition of W.

81. The system of claim 80, wherein the pitch periods are zero padded to N samples.

82. The system of claim 81, wherein a feature vector $\bar{u}_i$ is calculated as

$$\bar{u}_i = u_i \Sigma$$

where u_i is a row vector associated with a pitch period i, and Σ is the singular diagonal matrix.

83. The system of claim 82, wherein the distance between two feature vectors is determined by a metric comprising the cosine of the angle between the two feature vectors.

84. The system of claim 83, wherein the metric comprises a closeness measure, C, between two feature vectors, $\bar{u}_k$ and $\bar{u}_l$, wherein C is calculated as

$$C(\bar{u}_k, \bar{u}_l) = \cos(u_k\Sigma, u_l\Sigma) = \frac{u_k \Sigma^2 u_l^T}{\| u_k \Sigma \| \, \| u_l \Sigma \|}$$

for any $1 \leq k, l \leq 2KM$.

85. The system of claim 84, wherein a difference $d(S_1,S_2)$ between two segments in the voice table, S_1 and S_2, is calculated as

$$d(S_1,S_2) = d_0(p_1, q_1) = 1 - C(\bar{u}_{p1}, \bar{u}_{q1})$$

where d_0 is the distance between pitch periods p_1 and q_1, p_1 is the last pitch period of S_1, q_1 is the first pitch period of S_2, $\bar{u}_{p1}$ is a feature vector associated with pitch period p_1, and $\bar{u}_{q1}$ is a feature vector associated with pitch period q_1.

86. The system of claim 85, wherein the calculation for the difference between two segments in the voice table, S_1 and S_2, is expanded to include a plurality of pitch periods from each segment.

87. The system of claim 85, wherein the difference between two segments in the voice table, S_1 and S_2, is associated with a discontinuity between S_1 and S_2.

88. The system of claim 84, wherein a difference $d(S_1,S_2)$ between two segments in the voice table, S_1 and S_2, is calculated as

$$d(S_1,S_2) = \left| \frac{d_0(p_1, q_1) - d_0(p_1, \bar{p}_1) + d_0(q_1, \bar{q}_1)}{2} \right| = \left| \frac{C(\bar{u}_{p1}, \bar{u}_{\bar{p}1}) + C(\bar{u}_{q1}, \bar{u}_{\bar{q}1}) - C(\bar{u}_{p1}, \bar{u}_{q1})}{2} \right|$$

where d_0 is the distance between pitch periods, p_1 is the last pitch period of S_1, $\bar{p}_1$ is the first pitch period of a segment contiguous to S_1, q_1 is the first pitch period of S_2, $\bar{q}_1$ is the last pitch period of a segment contiguous to S_2, $\bar{u}_{p1}$ is a feature vector associated with pitch period p_1, $\bar{u}_{q1}$ is a feature vector associated with pitch period q_1, $\bar{u}_{\bar{p}1}$ is a feature vector associated with pitch period $\bar{p}_1$, and $\bar{u}_{\bar{q}1}$ is a feature vector associated with pitch period $\bar{q}_1$.

89. The system of claim 74, wherein the process further causes the processing unit to associate the distance between the feature vectors with speech segments in the voice table.

90. The system of claim 89, wherein the process further causes the processing unit to select speech segments from the voice table based on the distance between the feature vectors.

91. The system of claim 77, wherein the portions include centered pitch periods.

92. The system of claim 91, wherein the matrix W is a $(2(K\text{-}1)+1)M \times N$ matrix represented by

$$W = U \Sigma V^T$$

where K-1 is the number of centered pitch periods near the segment boundary extracted from each segment, N is the maximum number of samples among the centered pitch periods, M is the number of segments in the voice table having a segment boundary within the phoneme, U is the $(2(K\text{-}1)+1)M \times R$ left singular matrix with row vectors u_i $(1 \leq i \leq (2(K\text{-}1)+1)M)$, Σ is the $R \times R$ diagonal matrix of singular values $s_1 \geq s_2 \geq \ldots \geq s_R > 0$, V is the $N \times R$ right singular matrix with row vectors v_j $(1 \leq j \leq N)$, $R \ll (2(K\text{-}1)+1)M)$, and ^{T} denotes matrix transposition, wherein decomposing the matrix W comprises performing a singular value decomposition of W.

93. The system of claim 92, wherein the centered pitch periods are symmetrically zero padded to N samples.

94. The system of claim 93, wherein a feature vector $\bar{u}_i$ is calculated as

$$\bar{u}_i = u_i \Sigma$$

where u_i is a row vector associated with a centered pitch period i, and Σ is the singular diagonal matrix.

95. The system of claim 94, wherein the distance between two feature vectors is determined by a metric comprising a closeness measure, C, between two feature vectors, $\bar{u}_k$ and $\bar{u}_l$, wherein C is calculated as

$$C(\bar{u}_k, \bar{u}_l) = \cos(u_k\Sigma, u_l\Sigma) = \frac{u_k \Sigma^2 u_l^{\,T}}{\left\| u_k \Sigma \right\| \left\| u_l \Sigma \right\|}$$

for any $1 \leq k, l \leq (2(K\text{-}1)+1)M$.

96. The system of claim 95, wherein a difference $d(S_1,S_2)$ between two segments in the voice table, S_1 and S_2, is calculated as

$$d(S_1,S_2) = C(u\pi_{-1}, u\delta_0) + C(u\delta_0, u\sigma_1) - C(u\pi_{-1}, u\pi_0) - C(u\sigma_0, u\sigma_1)$$

where $u\pi_{-1}$ is a feature vector associated with a centered pitch period π_{-1}, $u\delta_0$ is a feature vector associated with a centered pitch period δ_0, $u\sigma_1$ is a feature vector associated with a centered pitch period σ_1, $u\pi_0$ is a feature vector associated with a centered pitch period π_0, and $u\sigma_0$ is a feature vector associated with a centered pitch period σ_0.

97. A machine-implemented method comprising:

gathering time-domain samples from recorded speech segments;

extracting features that represent the samples;

determining a discontinuity between the segments, the discontinuity based on a distance between the features.

98. The machine-implemented method of claim 97, wherein the time-domain samples include pitch periods surrounding a boundary of a phoneme.

99. The machine-implemented method of claim 98, wherein the features incorporate phase information of the pitch periods.

100. The machine-implemented method of claim 99, wherein extracting features comprises constructing a matrix from the time-domain samples and decomposing the matrix.

101. A machine-readable medium having instructions to cause a machine to perform a machine-implemented method comprising:

gathering time-domain samples from recorded speech segments;

extracting features that represent the samples;

determining a discontinuity between the segments, the discontinuity based on a distance between the features.

102. The machine-readable medium of claim 101, wherein the time-domain samples include pitch periods surrounding a boundary of a phoneme.

103. The machine-readable medium of claim 102, wherein the features incorporate phase information of the pitch periods.

104. The machine-readable medium of claim 103, wherein extracting features comprises constructing a matrix from the time-domain samples and decomposing the matrix.

105. An apparatus comprising:

means for gathering time-domain samples from recorded speech segments;

means for extracting features that represent the samples;

means for determining a discontinuity between the segments, the discontinuity based on a distance between the features.

106. The apparatus of claim 105, wherein the time-domain samples include pitch periods surrounding a boundary of a phoneme.

107. The apparatus of claim 106, wherein the features incorporate phase information of the pitch periods.

108. The apparatus of claim 107, wherein the means for extracting features comprises means for constructing a matrix from the time-domain samples and means for decomposing the matrix.

109. A system comprising:

a processing unit coupled to a memory through a bus; and

a process executed from the memory by the processing unit to cause the processing unit to gather time-domain samples from recorded speech segments, extract features that represent the samples, and determine a discontinuity between the segments, the discontinuity based on a distance between the features.

110. The system of claim 109, wherein the time-domain samples include pitch periods surrounding a boundary of a phoneme.

111. The system of claim 110, wherein the features incorporate phase information of the pitch periods.

112. The system of claim 111, wherein the process further causes the processing unit, when extracting features, to construct a matrix from the time-domain samples and decompose the matrix.